AD-A074 698   FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS   F/G 12/1
                DISTRIBUTIONAL RESULTS FOR RANDOM FUNCTIONALS OF A DIRICHLET PR--ETC(U)
                AUG 79   R C HANNUM, M HOLLANDER, N A LANGBERG   DAAG29-79-C-0158
UNCLASSIFIED   FSU STATISTICS-M520   ARO-16713.1-M   NL
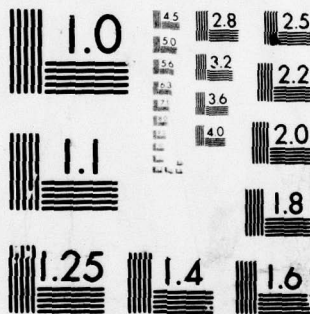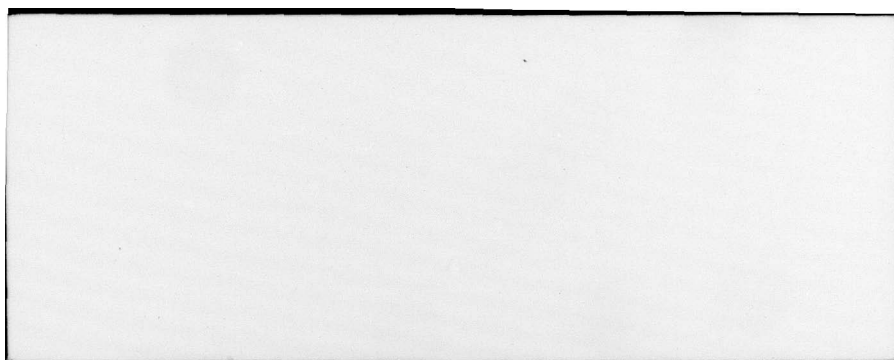
| OF |
AD
A074698

END
DATE
FILMED
11-79
DDC

1.0

4.5
5.0
5.6
6.3

2.8

2.5

1.1

3.2

3.6

4.0

2.2

2.0

1.25

1.4

1.8

1.6

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

(6)

DISTRIBUTIONAL RESULTS FOR RANDOM FUNCTIONALS OF A DIRICHLET PROCESS

BY

(10) Robert C. Hannum[1,2], Myles Hollander[1], and Naftali A. Langberg[3]

AMS 1970 subject classifications. Primary 60K99; Secondary 60E05.

Key words and Phrases. Dirichlet process, distribution of random functional.

SUMMARY

DISTRIBUTIONAL RESULTS FOR RANDOM FUNCTIONALS
OF A DIRICHLET PROCESS

is obtained

We obtain an expression for the distribution function of the random variable $\int Z\,dP$ where P is a random distribution function chosen by Ferguson's (1973) Dirichlet process on (R, B) (R is the real line and B is the σ-field of Borel sets) with parameter α, and Z is a real-valued measurable function defined on (R, B) satisfying $\int |Z|\,d\alpha < \infty$. As a consequence, we show that when α is symmetric about 0 and Z is an odd function, then the distribution of $\int Z\,dP$ is symmetric about 0.

alpha

infinity

the definite integral of the absolute value of Z to the derivative of alpha

the definite integral of ZdP

DISTRIBUTIONAL RESULTS FOR RANDOM FUNCTIONALS
OF A DIRICHLET PROCESS

by

Robert C. Hannum[1,2], Myles Hollander[2], and Naftali A. Langberg[3]
University of Denver and Florida State University

1. **Introduction and summary**. Let P be a random probability measure chosen by Ferguson's (1973) Dirichlet process on (R, B) (R is the real line and B is the σ-field of Borel sets) with parameter α. Let Z be a real-valued measurable function defined on (R, B) satisfying $\int |Z| d\alpha < \infty$. Ferguson's fundamental paper contains many results including, for the purpose of developing Bayesian nonparametric estimators of various parameters, results giving the mean of such random variables (rv's) as $\int Z dP$ and (variance P) and the median of the rv (median P). Yamato (1977) extends Ferguson's results by obtaining the moments of various random estimable parameters. Results pertaining to the distribution functions (df's) of random functionals are less available, an exception being Ferguson's (1973) expression for the df of the random q^{th} percentile of P.

In this note we study the df of $\int Z dP$. Our contributions are:

AMS 1970 *subject classifications*. Primary 60K99; Secondary 60E05.

Key words and phrases. Dirichlet process, distribution of random functional.

(i) In Theorem 2.5 we establish the equality

$$(1.1) \qquad \Pr\{\int Z dP \leq x\} = \Pr\{T^x \leq 0\},$$

which relates the df of $\int Z dP$ to the distributions of rv's T^x, $x \in R$, defined by (2.3) and Theorem 2.3. In Theorem 2.3 we obtain the characteristic function (cf) of T^x.

(ii) The cf of T^x can be used to obtain results concerning the distribution of $\int Z dP$. For example, in Corollary 2.6 we use the cf of T^x to prove that when Z is odd and α is symmetric about 0, then the distribution of $\int Z dP$ is symmetric about 0. (We also present an alternative proof of Corollary 2.6 using the gamma process representation (see Ferguson, 1973; Ferguson and Klass, 1972) of the Dirichlet process.) Thus, for example, when α is symmetric about 0, the odd moments of the random P, namely $\int \xi^n dP(\xi)$, $n = 1, 3, 5, \ldots$, have distributions that are symmetric about 0.

2. <u>The distribution of $\int Z dP$</u>. We assume the reader is familiar with definitions and basic results concerning Ferguson's Dirichlet process. If this is not the case, the reader is referred to Ferguson (1973). Throughout this section we take P to be a Dirichlet process on (R, B) with parameter α, and Z to be a real-valued measurable function defined on (R, B) satisfying $\int |Z| d\alpha < \infty$. We first prove two lemmas which will be used to show that the rv's T^x, $x \in R$, defined by (2.3) and Theorem 2.3, have continuous df's.

LEMMA 2.1. Let r be a positive integer and let $X_1, \ldots, X_r$ be independent and identically distributed (iid) rv's such that $\sum_{i=1}^{r} X_i$ has a continuous df. Then X_1 has a continuous df.

PROOF. Note that for all $y \in R$,

(2.1) $$(\Pr\{X_1 = y\})^r \leq \Pr\{\sum_{i=1}^{r} X_i = ry\}. \;||$$

LEMMA 2.2. Let $x \in R$ and let α be a finite measure on (R, B) such that $\alpha\{\xi: \; Z(\xi) \neq x\} > 0$, where Z is a measurable function integrable with respect to α. Then there is a positive integer r such that

$$\int_0^\infty \exp\{-r \int_{-\infty}^\infty \ln[1 + t^2\{Z(\xi) - x\}^2] d\alpha(\xi)dt < \infty.$$

PROOF. First note that since $\alpha\{\xi: \; Z(\xi) \neq x\} > 0$ then there exists two positive integers j and r such that $r \cdot \alpha\{\xi: \; |Z(\xi) - x| \geq j^{-1}\} > 1$. Letting $B_j = \{\xi: \; |Z(\xi) - x| \geq j^{-1}\}$ we then note that

$$\int_{-\infty}^\infty \ln[1 + t^2\{Z(\xi) - x\}^2]d\alpha(\xi) \geq \int_{B_j} \ln[1 + t^2\{Z(\xi) - x\}^2]d\alpha(\xi)$$

$$\geq \alpha(B_j) \cdot \ln[1 + (t/j)^2],$$

and so we have

$$\int_0^\infty \exp\{-r\int_{-\infty}^\infty \ln[1 + t^2\{Z(\xi) - x\}^2]d\alpha(\xi)\}dt \leq \int_0^\infty \exp\{-r \cdot \alpha(B_j) \cdot \ln[1 + (t/j)^2]\}dt$$

$$\leq \int_0^\infty [1 + (t/j)^2]^{-r \cdot \alpha(B_j)}dt \leq \int_0^\infty [1 + (t/j)^2]^{-1}dt < \infty. \;||$$

Now we introduce some notation. Let

$$I_{jk} = \begin{cases} \{\xi: \; (j - 1)/k < Z(\xi) \leq j/k\}, & \text{for } j = -k^2 + 1, \ldots, k^2 \\ \{\xi: \; Z(\xi) \leq -k\}, & \text{for } j = -k^2 \\ \{\xi: \; Z(\xi) > k\}, & \text{for } j = k^2 + 1. \end{cases}$$

Further let W_{jk}, $j = -k^2, \ldots, k^2 + 1$, be independent gamma rv's with shape parameters $\alpha(I_{jk})$ respectively, and common scale parameter 1. Let $x \in R$ and define, for $k = 1, 2, \ldots$

$$(2.2) \qquad T_k^x = \sum_{j=-k^2}^{-1} \{(j/k) - x\} W_{jk} + \sum_{j=0}^{k^2+1} \{[(j-1)/k] - x\} W_{jk}.$$

THEOREM 2.3. Let $x \in R$. Then the sequence T_k^x, $k = 1, 2, \ldots$, converges in law as $k \to \infty$ to a rv T^x whose cf is given by

$$(2.3) \qquad \phi_{T^x}(t) = \exp\{-\int_{-\infty}^{\infty} \ell n[1 - it\{Z(\xi) - x\}] d\alpha(\xi)\}.$$

PROOF. Let $\delta_\xi(I_{jk}) = 1$ when $\xi \in I_{jk}$ and 0 otherwise, and for each $\xi \in R$ define

$$Z_k(\xi) = \sum_{j=-k^2}^{-1} [j/k] \delta_\xi(I_{jk}) + \sum_{j=0}^{k^2+1} [(j-1)/k] \delta_\xi(I_{jk})$$

for $k = 1, 2, \ldots$. Let $x \in R$ and note from (2.2) that since T_k^x is a linear combination of independent gamma rv's, it has a cf which is infinitely divisible. Hence we may write

$$\ell n \, E(\exp\{it T_k^x\}) = \sum_{j=-k^2}^{-1} \ell n \, E(\exp\{it[(j/k) - x] W_{jk}\}) +$$

$$+ \sum_{j=0}^{k^2+1} \ell n \, E(\exp\{it[\{(j-1)/k\} - x] W_{jk}\})$$

$$= \sum_{j=-k^2}^{-1} \ell n[1 - it\{(j/k) - x\}]^{-\alpha(I_{jk})} + \sum_{j=0}^{k^2+1} \ell n[1 - it\{[(j-1)/k] - x\}]^{-\alpha(I_{jk})}$$

$$= -\{ \sum_{j=-k^2}^{-1} \alpha(I_{jk}) \ell n[1 - it\{(j/k) - x\}] + \sum_{j=0}^{k^2+1} \alpha(I_{jk}) \ell n[1 - it\{[(j-1)/k] - x\}]\}$$

$$= -\int_{-\infty}^{\infty} \ell n[1 - it\{Z_k(\xi) - x\}]d\alpha(\xi)$$

$$= -\int_{-\infty}^{\infty} \ell n[(1 + t^2\{Z_k(\xi) - x\}^2)^{\frac{1}{2}}]d\alpha(\xi) + i \int_{-\infty}^{\infty} \tan^{-1}[t\{Z_k(\xi) - x\}]d\alpha(\xi).$$

Now observe that $|Z_k(\xi)| \leq |Z(\xi)|$ for all real ξ and so we have

$$(2.5) \quad \ell n[(1 + t^2\{Z_k(\xi) - x\}^2)^{\frac{1}{2}}] \leq |t| \cdot \{|Z_k(\xi)| + x\} \leq |t| \cdot \{|Z(\xi)| + x\}$$

and

$$(2.6) \quad |\tan^{-1}[t\{Z_k(\xi) - x\}]| \leq |t| \cdot \{|Z_k(\xi)| + x\} \leq |t| \cdot \{|Z(\xi)| + x\}.$$

Since the right-hand-sides of (2.5) and (2.6) are integrable with respect to α and since $\lim_{k\to\infty} Z_k(\xi) = Z(\xi)$, then by the dominated convergence theorem we have from (2.4) that

$$(2.7) \quad \lim_{k\to\infty} E(\exp\{itT_k^x\}) = \exp\{-\int_{-\infty}^{\infty} \ell n[1 - it\{Z(\xi) - x\}]d\alpha(\xi)\}.$$

Writing the right-hand-side of (2.7) as $\exp\{-[G(t, x) + iH(t, x)]\}$ where

$$G(t, x) = \frac{1}{2} \int_{-\infty}^{\infty} \ell n[1 + t^2\{Z(\xi) - x\}^2]d\alpha(\xi),$$

$$H(t, x) = -\int_{-\infty}^{\infty} \tan^{-1}[t\{Z(\xi) - x\}]d\alpha(\xi),$$

it is straightforward to show that for any real number x, both $|G(t, x) - G(0, x)|$ and $|H(t, x) - H(0, x)|$ are bounded above by $|t| \cdot \int_{-\infty}^{\infty} \{|Z(\xi)| + x\}d\alpha(\xi)$, which goes to zero as $t \to 0$. This shows that the right-hand-side of (2.7) is continuous at $t = 0$ and hence the result of the theorem follows from the continuity theorem for cf's. ||

LEMMA 2.4. Let $x \in R$ and assume that $\alpha\{\xi: Z(\xi) \neq x\} > 0$. Then T^x has a continuous df.

PROOF. By Lemma 2.2 there is a positive integer r such that $\int_{-\infty}^{\infty} |[E(\exp\{itT^x\})]^r| dt < \infty$. Now $[E(\exp\{itT^x\})]^r$ is the cf of the convolution of r iid rv's distributed as T^x. Since the cf of this convolution is absolutely integrable we conclude (cf. Breiman (1968), p. 178) that the convolution of r iid rv's distributed as T^x is continuous. Consequently the result follows from Lemma 2.1. $||$

We are now ready to obtain an expression for the df of $\int Z dP$.

THEOREM 2.5. Let P be a Dirichlet process on (R, B) with parameter α and let Z denote a real-valued measurable function defined on (R, B) integrable with respect to α.

(i) If there exists a real number x such that $\alpha\{\xi: Z(\xi) \neq x\} = 0$, then $\int Z dP$ is degenerate at x.

(ii) If $\alpha\{\xi: Z(\xi) \neq x\} > 0$ for all $x \in R$, then (1.1) holds, where T^x is the rv whose cf is given by the right-hand-side of (2.3).

PROOF. (i) Note that $\alpha\{\xi: Z(\xi) \neq x\} = 0$ implies that $P\{\xi: Z(\xi) \neq x\} = 0$. Thus $\int_{-\infty}^{\infty} Z(\xi) dP(\xi) = x$.

(ii) Let $A_k = \int_{-\infty}^{\infty} Z_k(\xi) dP(\xi)$, for $k = 1, 2, \ldots$, and note that

$$\Pr\{A_k \leq x\} = \Pr\{\int_{-\infty}^{\infty} Z_k(\xi) dP(\xi) \leq x\}$$

(2.8)

$$= \Pr\{\sum_{j=-k^2}^{-1} [j/k] P(I_{jk}) + \sum_{j=0}^{k^2+1} [(j-1)/k] P(I_{jk}) \leq x\}.$$

Now for each k the random vector $(P(I_{jk}))$, $j = -k^2, \ldots, k^2 + 1$, has a Dirichlet distribution with parameter $(\alpha(I_{jk}))$, $j = -k^2, \ldots, k^2 + 1$. Letting W_{jk}, $j = -k^2, \ldots, k^2 + 1$, represent independent gamma rv's with shape parameters $\alpha(I_{jk})$, $j = -k^2, \ldots, k^2 + 1$, respectively and common scale parameter 1, and using the gamma distribution definition of the Dirichlet distribution, (2.8) becomes

$$\Pr\{A_k \le x\} = \Pr\{\sum_{j=-k^2}^{-1} [j/k]W_{jk} + \sum_{j=0}^{k^2+1} [(j - 1)/k]W_{jk} \le x \sum_{j=-k^2}^{k^2+1} W_{jk}\}$$

$$(2.9) \qquad = \Pr\{\sum_{j=-k^2}^{-1} \{(j/k) - x\}W_{jk} + \sum_{j=0}^{k^2+1} \{[(j - 1)/k] - x\}W_{jk} \le 0\}$$

$$= \Pr\{T_k^x \le 0\}.$$

Now note that $\lim_{k \to \infty} A_k = \int ZdP$ almost everywhere with respect to α and thus by (2.9) we have

$$\Pr\{\int ZdP \le x\} = \lim_{k \to \infty} \Pr\{A_k \le x\} = \lim_{k \to \infty} \Pr\{T_k^x \le 0\} = \Pr\{T^x \le 0\},$$

the last equality following from Theorem 2.3 and Lemma 2.4. ||

Using Theorem 2.5 in conjunction with the cf of T^x, given in (2.3), it is now possible to gain additional information concerning the distribution of the random mean $\int ZdP$.

COROLLARY 2.6. Let P be a Dirichlet process on (R, B) with parameter α, where α is a positive finite and symmetric measure about the origin. Let Z denote a real-valued measurable and odd function such that $\int |Z|d\alpha < \infty$. Then $\int ZdP$ has a symmetric distribution about the origin.

PROOF. First assume that there is a real number x such that
$\alpha\{\xi: \ Z(\xi) \neq x\} = 0$. Then since $Z(\xi) = -Z(-\xi)$ it follows that $\alpha\{\xi: \ Z(\xi) \neq 0\} = 0$.
Thus by Theorem 2.5 $\int ZdP$ is degenerate at $x = 0$ and hence symmetric.

Next assume that $\alpha\{\xi: \ Z(\xi) \neq x\} > 0$ for all $x \in R$. By repeating our
arguments it is clear that $Pr\{\int ZdP \geq -x\} = Pr\{T^{-x} \geq 0\}$ and thus to show that
$\int ZdP$ is symmetric about the origin it suffices to show that, for all $x \in R$,

$$(2.10) \qquad\qquad Pr\{T^{x} \leq 0\} = Pr\{-T^{-x} \leq 0\}.$$

Since $Z(\xi) = -Z(-\xi)$ and since $d\alpha(\xi) = d\alpha(-\xi)$ we have

$$\ln E(\exp\{itT^{-x}\}) = -\int_{-\infty}^{\infty} \ln[1 - it\{Z(\xi) + x\}]d\alpha(\xi)$$

(2.11)

$$= \int_{-\infty}^{\infty} \ln[1 + it\{Z(-\xi) - x\}]d\alpha(-\xi) = \ln E(\exp\{-itT^{x}\}).$$

From (2.11) we obtain that T^{x} has the same df as $-T^{-x}$ and hence (2.10)
follows directly. ||

The proof of Corollary 2.6 given above is based on Ferguson's (1973)
original definition of the Dirichlet process. Below we exhibit an alternative
proof of Corollary 2.6 which is based on Ferguson's (1973) (also see Ferguson
and Klass, 1972) alternative definition of the Dirichlet process as the sum of
a countable number of jumps of random height at a countable number of random
points. (This alternative representation is often useful in obtaining results
for the Dirichlet process, cf. Ferguson (1973), Korwar and Hollander (1973),
Yamato (1977).)

ALTERNATIVE PROOF OF COROLLARY 2.6. We have, by Theorem 2 of Ferguson
(1973),

$$\int Z dP = \sum_{j=1}^{\infty} Z(V_j) P_j$$

where V_1, V_2, ... are iid according to $Q(\cdot) = \alpha(\cdot)/\alpha(R)$, where the P_j's are independent of the V_j's, and where the P_j's depend on $\alpha(\cdot)$ only through $\alpha(R)$. Thus we have (since Z is odd)

$$(2.12) \qquad \int Z dP = \sum_{j=1}^{\infty} Z(V_j) P_j = - \sum_{j=1}^{\infty} Z(-V_j) P_j.$$

Now since α is symmetric about the origin, $-V_j \overset{d}{=} V_j$, for all $j = 1, 2, \ldots,$ where "$\overset{d}{=}$" means "has the same distribution as." Hence, since the V_j's are independent of the P_j's we have

$$(2.13) \qquad - \sum_{j=1}^{\infty} Z(-V_j) P_j \overset{d}{=} - \sum_{j=1}^{\infty} Z(V_j) P_j = - \int Z dP.$$

From (2.12) and (2.13) we conclude that $\int Z dP \overset{d}{=} - \int Z dP$ and so the df of $\int Z dP$ is symmetric about the origin. ||

REFERENCES

[1] BREIMAN, L. (1968). Probability. Addison-Wesley, Reading, Mass.

[2] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. Ann. Statist. 1 209-230.

[3] FERGUSON, T. S. and KLASS, M. J. (1972). A representation of independent increment processes without Gaussian components. Ann. Math. Statist. 43 1634-1643.

[4] KORWAR, R. M. and HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. Ann. Prob. 1 705-711.

[5] YAMATO, H. (1977). Relation between limiting Bayes estimates and the U-statistics for estimable parameters of degrees 2 and 3. Commun. Statist. - Theor. Math. A6(1) 55-66.

Security Classification of this Page

REPORT DOCUMENTATION PAGE

1. REPORT NUMBERS	2. GOVT. ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
FSU No. 520 AFOSR 78-3678 No. 15 USARO-DAAG-29-79-C-0158 No. 38		

4. TITLE	5. TYPE OF REPORT & PERIOD COVERED
Distributional results for random functionals of a Dirichlet process	Technical Report
	6. PERFORMING ORG. REPORT NUMBER
	FSU Statistical Report No. M484

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Robert C. Hannum, Myles Hollander, and Naftali A. Langberg	AFOSR 78-3678 No. 15 USARO-DAAG-29-79-C-0158 No. 38

9. PERFORMING ORGANIZATION NAME & ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA AND WORK UNIT NOS.
The Florida State University Department of Statistics Tallahassee, Florida 32306	

11. CONTROLLING OFFICE NAME & ADDRESS	12. REPORT DATE
Air Force Office of Scientific Research F Bolling Air Force Base, D.C. 20332 United States Army Research Office Durham, North Carolina	August, 1979
	13. NUMBER OF PAGES
	9

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS (of this report)
	Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Dirichlet process, distribution of random functional

20. ABSTRACT

We obtain an expression for the distribution function of the random variable $\int Z dP$ where P is a random distribution function chosen by Ferguson's (1973) Dirichlet process on (R, B) (R is the real line and B is the α-field of Borel sets) with parameter α, and Z is a real-valued measurable function defined on (R, B) satisfying $\int |Z| d\alpha < \infty$. As a consequence, we show that when α is symmetric about 0 and Z is odd ufunction, then the distribution of $\int Z dP$ is symmetric about 0.